

# 算法治理中的算法公共性及其实现

## Publicity of Algorithms and Its Realization in Algorithmic Governance

张海柱 /ZHANG Haizhu

(东北大学文法学院, 辽宁沈阳, 110169)  
(School of Humanities and Law, Northeastern University, Shenyang, Liaoning, 110169)

**摘要:** 随着公共事务治理场景中算法工具的广泛应用, 算法治理逐渐成为公共管理或治理的新形态。算法治理以公共利益或公共价值的实现为根本目标。而基于技术政治视角的分析揭示了算法的政治性, 它不仅可以被应用于实现特定政治意图, 其本身的技术特性也会与特定政治价值相关联并产生社会政治影响。因此算法自身的公共性应当成为算法治理的重要基础。算法公共性包括算法的公益性、公开性与公平性三个方面, 它们在现实治理实践中面临着异化风险。算法公共性的实现需要一种参与式设计的新理念, 它强调通过算法设计过程中多元主体的广泛参与来实现社会整体价值的嵌入与算法技术发展方向的调整, 从而使算法的技术属性更好地契合公共性价值。

**关键词:** 算法治理 算法公共性 技术政治 参与式设计

**Abstract:** With the widespread use of algorithmic tools in the context of public affairs governance, algorithmic governance is gradually becoming a new form of public management or governance. The fundamental goal of algorithmic governance is the realization of public interest or public values. From a techno-political perspective, this paper reveals the political nature of algorithms, which can be applied to achieve specific political intentions. Moreover, their own technical characteristics can also be associated with specific political values and have socio-political implications. Thus, the publicity of the algorithm itself should be an important basis for algorithmic governance. The publicity of algorithms includes the public good, openness and fairness of algorithms, which face the risk of alienation in real governance practices. The realization of algorithm publicity requires a new idea of participatory design, which emphasizes the embedding of overall social values and adjustment of algorithm technology development direction through the extensive participation of multiple subjects in the algorithm design process. This approach will enable the technical properties of algorithms can better fit the public values.

**Key Words:** Algorithmic governance; Publicity of algorithms; Techno-politics; Participatory design

中图分类号: B82-05; TP301.6 文献标识码: A DOI: 10.15994/j.1000-0763.2023.06.003

## 引言

近年来, 以机器学习为代表的算法研发取得了重要突破, 成为新一代人工智能技术的核心构成。作为一种计算程序或规则, 算法因其

高效的大数据处理和精准的预测分析能力而被广泛用于辅助决策甚至自动化决策, 其应用范围也由私人商业领域逐渐向司法裁决、社会治理或政府管理等公共事务治理场景拓展, 由此产生了“算法治理”(algorithmic governance)这一公共管理或治理的新形态甚至新范式。<sup>[1]</sup>

基金项目: 国家社会科学基金一般项目“人类命运共同体的传统思想来源研究”(项目编号: 21BZZ037)。

收稿日期: 2023年2月7日

作者简介: 张海柱(1985-)男, 河北盐山人, 东北大学文法学院教授, 研究方向为政策过程、风险治理与数字治理。Email: haizhu1213@sina.com

与私人部门应用算法工具对私人或商业利益的追求不同，公共部门的算法治理应以公共利益的实现或公共价值的创造为根本目标和原则，<sup>[2]</sup>由此公共性应当成为算法治理的基本属性或价值原则<sup>①</sup>。

然而，算法应用可能出现侵犯个人隐私以及算法偏见、歧视、操纵等风险，机器学习算法的不透明或不可解释性则使得算法决策面临违反正当程序、难以进行有意义的监督与问责等问题，这些均对算法治理公共性的实现构成了直接挑战。此外，由于政府公共部门普遍缺乏算法技术研发的资源与专业能力，因此多通过政府采购或外包的方式引入科技企业研发的算法系统，甚至将系统的整体设计、建设、运营与维护均交由私人部门。<sup>[3]</sup>这使得算法治理日益呈现出“混合”治理的形态，私人部门不断介入甚至主导算法治理的过程与结果。由于私人与公共部门之间在目标和价值取向上存在显著差异甚至冲突，因此如何防范算法治理的公共性被商业资本的私利性所替代或异化，成为有待解决的难题。

针对上述问题，既有研究从算法治理的主体重塑（例如塑造算法平台的公共性<sup>[4]</sup>）、伦理嵌入（例如将公平、负责任等伦理价值嵌入算法规则<sup>[5]</sup>）以及过程控制（例如对算法应用过程进行监督、审计和问责<sup>[6]</sup>）等角度提出了应对思路。然而较少有研究者关注作为治理工具的“算法”本身与公共性之间的关系。之所以如此，是因为许多人在有意识或无意识中持有一种技术或工具的“中立性”思维，认为算法应用结果主要取决于设计者、控制者和使用者的价值立场、行为意图及使用方式。然而在科技与社会研究（STS）领域，“技术物”（artifacts）的政治性<sup>[7]</sup>命题提醒我们技术不仅会在设计与应用过程中被人为地嵌入政治意图，而且其本身内在地具有政治性，会产生特定的社会政治影响。因此，我们需要认真审视算法这一“技术物”的政治性，这使“算法公

共性”议题的提出成为可能。

有鉴于此，本文将借鉴技术政治相关理论来对算法的政治性属性进行解析，进而基于公共性理论来讨论算法公共性的价值构成及其异化风险。本文最后将提出一种参与式算法设计的思路，以期在算法治理中实现算法公共性价值的实现提供启发借鉴。

## 一、算法公共性议题的提出： 技术政治视角的分析

对技术“中立性”幻象的批判是当代技术哲学以及STS领域研究的重要主题。例如加拿大学者斯迈思（Dallas Smythe）在其经典的“自行车之后是什么？”一文中指出了技术的政治与意识形态属性，认为不同的技术会服务于不同的政治社会制度，并且一针见血地指出技术中立观点“本身就是具有政治性的，而且是反动的”。<sup>[8]</sup>美国学者温纳（Langdon Winner）在“技术物有政治性吗？”一文中更为细致地讨论了技术或“技术物”的政治属性问题。<sup>[7]</sup>在温纳看来，技术物会以两种方式呈现政治性或产生政治影响。第一种是技术物“外在”的政治性，即技术物在特定情形下经由特定形式的设计、配置或安排而服务于特定的政治意图或产生特定的政治后果。此时，技术物不再只是中立的特定功能的提供者，而成为政治意图的媒介，可能服务于特殊利益或价值。

更加值得注意的是第二种形式的技术政治性，即技术物“内在”的政治性。这种政治性与行为主体的意图和使用方式无关，而是由于技术物自身的特性导致其与特定政治价值之间存在必然联系。温纳指出，“选择特定技术即是选择特定形式的政治生活”。<sup>[7]</sup>他以核电为例，认为该技术倾向于将社会导向威权主义的管理模式，而分散式的太阳能技术则与民主社会更为兼容。也即，这种内在政治性反映了技术物自身的“能动性”。当特定技术被引入社

<sup>①</sup>现实中，私人性质的平台巨头企业的算法应用产生了广泛的社会乃至政治影响，其中涉及的公共性议题值得探讨，不过这不在本文的研究范围之内。

会系统后,社会系统中的人、组织以及制度安排势必会或多或少地做出调整以进行“配合”,而这种调整可能涉及利益与价值的重新分配,从而体现为一种政治性影响。

基于上述技术政治视角来审视“算法”这一技术物,其外在政治性主要是指特定主体(算法设计者、控制者或使用者)基于特定意图、价值目标或伦理原则而设计和使用算法,从而产生各种积极或消极的社会政治后果。这一点已经得到了较多讨论和揭示,例如算法背后可能体现的是商业逻辑或资本逐利性,算法设计者也可能故意选择某些具有伦理属性的参数设置(种族、性别等)从而导致社会歧视性后果。然而对于算法的内在政治性,则较少受到关注或得到承认。例如有研究明确指出“从技术层面而言,算法是中性的”,“算法只是公共治理主体手中的一把枪,究竟如何使用这把枪,最终还是取决于公共治理主体”。<sup>[9]</sup>

如前所述,算法的内在政治性与人的意图、价值取向或使用方式无关,而主要取决于算法自身的技术特性。由此我们需要认真审视算法的技术逻辑及其相应的社会政治影响潜力。简单来说,算法在基本技术形态上主要体现为基于“If···Then”条件判断式的代码以及由此构成的复杂模型、程序或系统,算法的底层技术逻辑即体现在各种“如果”(if)条件(例如资料特征是否满足模型参数及其阈值)和“结果”(then)选项(例如对资料执行过滤、选取、分类、排序等操作)的设置、组合与选择中。以此为基础,算法的设计与实践运行中呈现出了一系列独特的技术特征,包括数据依赖性、统计相关性以及相对自主性等,并由此可能造成特定的社会政治后果。

首先,数字化或数据化是算法运行的基础和前提,能够通过各种数字技术手段全面准确地实现包括人类自身在内的现实世界向数字世界的转化是算法的潜在预设。然而数字世界并非现实世界的准确映射,许多人、物、价值或目标因无法数据化而被忽视或遗弃,从而造成“非存在的社会存在”(non-existent existence)<sup>[10]</sup>以及数据中的历史偏见问题。同时对大量

初始训练数据和用户反馈数据的依赖也更有利于大型平台的算法开发,使其走向垄断。其次,算法体现的是数据资料主导而非规则主导的逻辑,它的输出结果呈现的是数据统计上的相关性而非因果关系。由此导致算法的结果难以进行解释和监督、问责,而算法运行中统计性偏见(statistical biases)的存在也会导致歧视性后果。<sup>[11]</sup>此外,机器学习算法与其他技术物的最大区别在于它具有一定的自主性,即可以在大量数据中自动提炼特征、建立关联并生成模型,并随着数据更新而自我更迭。正是这一特征为人工智能的实现提供了巨大的想象空间。然而这种自主性也意味着算法结果的不确定性,可能产生“非预期后果”甚至失控,造成严重社会危害。

由上可知,由于算法的技术特征,即便是没有特殊利益或价值意图的嵌入以及算法使用中的故意错误或失误操作,算法自身也内含政治性并会导致各种社会政治后果,在优先满足特定群体利益或价值的同时,对其他群体造成伤害,从而导致对社会公平、公共利益等公共性价值的偏离。因此,当我们以公共性价值来审视公共事务场景中算法治理的合理性或正当性时,除了对治理主体的目标、价值取向以及治理过程和结果进行考察外,还需要关注作为治理工具的算法自身的政治与价值属性。算法作为“技术物”具有一定的可塑性,这种可塑性使得通过改变算法设计的逻辑或技术构成可以使算法与不同的价值原则相契合。那么,是否可以通过技术设计的优化而使算法自身能够更好地与公共性价值产生内在关联?正是基于这种考虑,本文提出“算法公共性”这一研究议题,这将为算法治理中公共性价值的实现提供新的思考方向。

## 二、算法治理中的算法公共性: 价值构成及异化风险

算法治理是政府等公共部门基于算法技术手段的辅助来解决公共问题、提供公共服务或追求公共利益与价值的实践过程。随着公共治



理场景中算法重要性的日益提升,以及政府官僚机构对算法工具依赖性的不断增强,算法的公共基础设施属性也在不断凸显。<sup>[12]</sup>而当算法成为公共治理的重要基础设施之后,意味着它必须能够有效地服务于或支撑公共治理主体对于公共性价值的追求。当然,前文对算法政治性的分析表明,算法对公共性价值的“服务”或“支撑”不仅仅是指它可以被特定主体用于追求公共价值这一“政治意图”,还是指算法自身的技术特征能够与公共治理的制度体系及其价值原则相容,进而产生积极的社会政治后果。正是在此意义上,本文提出了“算法公共性”这一研究议题,它对于算法治理场景中公共性价值的实现有着不可忽视的重要作用<sup>①</sup>。

为了更好地理解“算法公共性”的内涵与价值定位,还需要对“公共性”本身进行解析。一般而言,公共性是与私人性、隐秘性、封闭性或排他性等相对而言的概念,体现出“共有性、包容性和公开性”等特点。<sup>[13]</sup>同时,作为一个在政治哲学、公共行政等学科广泛讨论的概念,公共性也被赋予了多样的内涵。例如阿伦特、哈贝马斯等学者侧重对公共领域的公共性的考察,由此公开可见、可进入、理性交往、协商对话被视为公共性的重要内涵。<sup>[14]</sup>而公共行政领域学者更多是通过公共利益与公共物品两个方面来界定公共行政的公共性。<sup>[15]</sup>此外,无论是罗尔斯等政治学家对分配正义的考察,还是弗雷德里克森等公共行政学者对公平正义价值观的讨论,社会公平或公正始终是公共性的重要内涵。<sup>[14]</sup>本文认为,上述价值在某种程度上均应当成为算法治理中算法公共性的内在构成,而这又可以进一步归纳为公益性、公开性与公平性三个方面。下文将分别从这三个方面入手,对算法公共性的价值构成及其在治理实践中面临的异化风险进行分析。

### 1. 算法公益性及其异化风险

算法公益性是指公共治理实践中算法工具的设计与使用应当以公共利益的改善为目标,

能够服务于多数人而非少数人的利益。然而,长期以来算法主要是基于商业场景开发设计,因此主要服务于商业利益或资本利益,并形成了大型科技公司或平台巨头企业主导算法研发的局面。政府等公共部门由于缺乏独立开发算法系统的资源与能力,往往通过政府购买的方式引入由私人企业研发的算法系统。然而,治理场景的转换并不会使算法工具自然而然地转变性质。正如有学者指出的,在平台经济模式下人类已经成为一种“资源”,在算法治理中亦是如此,“与公共利益相反,源自平台经济的算法治理的私人性质自然优先将人类视为一种资源”。([16], p.181)这种将人视为客体化的资源而非具有主体性的服务对象的倾向,成为算法治理中算法公益性异化的重要风险。

同时由于算法系统的复杂性,私人科技公司往往会全面介入其设计、部署、运行与日常维护。在此过程中,掌握着算法技术的企业及其背后的资本力量将化身为公共治理过程中的“影子官僚”。<sup>[9]</sup>逐利性是算法影子官僚的根本特征,如何克制其逐利冲动很大程度上取决于政府部门的监管能力。然而由于政府与科技公司之间明显的信息与技术能力的不对称,这种监管难以取得实效。现实结果是算法研发者在商业逻辑主导下“更加倾向于用有利于其利润最大化的算法代码与算法程序去影响甚至支配政府的日常行为,进而演化成为算法权力”,<sup>[9]</sup>从而以私人利益置换公共利益。

而当算法被引入公共治理过程后,政府公共部门也会为了“配合”算法的应用而进行相应的行为调整,这正是算法政治性的重要体现。这方面的重要体现是越来越多政府部门允许甚至依赖私人科技公司来处理公共数据。同样,在商业或资本逻辑的驱动下,私人企业对公共数据的占有和处理可能引发侵犯公民隐私权利、危害社会秩序甚至国家安全的风险。而机器学习算法所具有的“在数据中发现意料之外的模式与为全新用途而调整数据使用的能力”

<sup>①</sup>在算法治理中公共性价值的实现上,算法的“外在”政治性与“内在”政治性之间的关系有待进一步考察。也即,作为技术物的算法自身的公共性对于算法治理中公共性价值的实现而言,是否是不可或缺的,这一点还需要谨慎讨论。感谢审稿专家对这一问题的提醒。

意味着科技公司对公共数据的“合法”使用也可能引发不可预知的风险后果。( [16], p.22 )

算法的引入造成政府行为调整的另一种体现是可能引发官僚的责任规避倾向。也即,如果官僚能够证明其决策是基于算法系统的建议,就可以免于承担决策失误的责任,这正是许多官僚机构及其工作人员乐于使用算法工具的重要原因。<sup>[17]</sup>然而如果对算法的采用主要是基于“避责”考虑的话,事实上也意味着个人或部门利益(避责)对公共利益追求的置换。

## 2. 算法公开性及其异化风险

算法公开性是指算法治理中算法工具的“可见性”,要能够以恰当方式进行公开呈现,并提供充分的机会让利益相关主体针对算法的设计和讨论进行监督。这种公开性有着多方面的内涵与要求。首先,算法公开性是指算法“存在”意义上的公开,即政府等公共治理主体要明确公开说明其治理决策是否以及在何时使用了算法工具。然而现实中算法的运行具有较强的隐蔽性,难以从治理结果上区分是否使用了算法工具,这使得人们往往在无意识中便成为了算法计算的对象。这种“算法无意识”状态与算法的不公开之间相互强化,容易演变为难以打破的恶性循环。

其次,算法公开性是指算法“关系”意义上的公开,即要明确公开算法的设计者、所有者、控制者和使用者等主体构成及其利益关联。如前所述,在算法治理过程中,作为算法使用者的政府部门与作为算法设计者的科技公司之间存在明显的信息和技术不对称,后者往往成为算法的实际控制者,并可能通过隐蔽的方式以商业利益来置换公共利益。正如有学者曾指出的,明确算法如何工作固然重要,但“算法为谁工作”这一问题更加不容忽视。<sup>[18]</sup>当前对于算法公开或透明的关注主要聚焦于算法模型本身,在一定程度上忽视了算法背后的真实利益主体及其关联问题,这构成了算法公开性异化的重要风险。

此外,算法公开性还涉及算法“构成”意义上的公开,即要明确公开算法的基本原理、目标、意图和嵌入的价值标准,以及算法的初

始训练数据和最终输入数据的来源及构成。同时,政府部门还应当基于“正当程序”的要求来对算法决策的合理性与正当性进行公开解释或说明理由。不过,现实中基于保护商业或国家秘密的考虑以及机器学习算法运行原理上的难以解释(甚至算法设计者也无法解释)等原因,算法往往被视为不透明的“黑箱”,<sup>[19]</sup>并由此成为无法公开的理由。然而事实上,对上述算法构成要素的公开以及正当程序意义上的说明理由并不要求完全披露算法源代码或精通算法原理,与商业或国家秘密的保护原则也不会产生必然的冲突(例如公开算法的目标或功能而非代码)。<sup>[20]</sup>就此而言,因算法“黑箱”的存在而主张“无法公开”在很大程度上即成为一种有意规避公开的托词。

## 3. 算法公平性及其异化风险

算法公平性是指算法工具的设计和应用应当有助于公共治理主体对于社会公平这一重要价值目标的追求,防止算法偏见、歧视等不公平现象的出现。公共治理实践应当追求公平价值这一点毋庸置疑,然而作为治理工具的算法本身是否能够体现公平性或者与公平价值相容,则有待进一步考察。事实上,正如有学者曾指出的,人类在研发某种技术物时所追求的目标,与技术物本身的特性密切相关。<sup>[21]</sup>就算法技术而言,其重要特性之一在于处理大数据上的高效率 and 准确性。然而在算法治理中,对于效率和准确性的优先强调可能造成对公平等社会价值的损害。例如,如果算法的输入数据中隐含着历史偏见,那么高效准确的算法会“被用来让已经存在歧视的系统运行得更顺畅”,([22], p.114)从而会固化甚至扩大既有的偏见等社会不公平状况。

同时,由于机器学习算法运行是基于统计相关性而非因果关系,因此很难对算法的结果进行传统意义上的解释,当出现错误结果时也很难进行有效的原因追溯。更有甚者,算法的运行可能出现难以发现和纠正的负反馈循环问题。例如在警务算法的运用中,如果算法错误地将某个地区标注为“高犯罪风险”,那么由此导致的警力投入的增加会带来更高的逮捕



率，而这一结果将作为警务算法新的训练数据来强化原有的预测模型。此时，看似“准确”的算法预测事实上仅仅只是一种“自我实现的预言”，<sup>[23]</sup>具有明显的虚假性和误导性。

此外，机器学习算法在处理数据特别是大型非结构化数据集上的高效率使其能够确定人类认知难以辨识的隐含模式和相关性，加上机器学习算法所具有相对自主性特征，这些均可能导致算法的运行结果偏离设计者的初始目标和意图，从而产生非预期后果。例如，为了确保算法结果的公平，算法设计者在对资料进行特征提取和标签时，会选择主动屏蔽性别、种族等敏感特征。然而，这种看似公平的做法有时却难以奏效，因为算法强大的数据关联与分析能力往往使其能够在未被屏蔽的非敏感特征信息中识别出敏感性特征要素。例如在一些国家，特定种族的居民往往聚集生活在相对固定的邻近区域，此时相关算法工具即便屏蔽了种族特征，但却可以通过家庭位置等信息自动识别出其种族属性，从而产生歧视性的意外后果。这种情况下的算法不公平问题具有极强的隐蔽性，难以揭示和纠正。

### 三、算法公共性的实现： 一种参与式设计理念

在技术哲学与STS领域研究中，面对技术政治性的彰显而引发的负面社会政治后果乃至技术失控的可能性，诉诸于技术的民主干预或控制被视为重要的应对思路。该种思路“试图通过公共对话和协商民主来激发技术领域的民主潜能，实现技术权力调节与专业知识生产的民主化”，<sup>[24]</sup>从而打造出更为适当的技术。正如温纳在对现代技术失控问题的反思中明确指出的，“现代技术的一个重要缺陷在于，那些因它们的存在而受到影响的人们对于其设计和运行只有很少的或没有控制权”，“因此，涉及技术的规划、建造和控制过程应该尽可能向那些注定要体验最终产品和全面社会后果的人敞开”，（[25]，p.280）这正是技术公共性的直接体现。

上述技术民主路径为算法治理中算法公共性的实现提供了重要启发。长期以来，算法被视为技术理性或工具理性发展到极致的体现，算法的开发、设计与应用则被视为纯粹且高端的“技术”领域，不仅公众等利益相关者无法参与，甚至政府部门等算法系统的采购者和使用者也在很大程度上丧失了控制权。正如有学者所指出的，“在引入算法后，算法影子官僚把本应由政府部门、公众和专家等共同参与、互动协商的执行过程，退化为算法公司秘密进行的算法设计和数据处理过程，把决策责任交给了充斥复杂计算与最优化的模型和冰冷的数据”。<sup>[9]</sup>

算法治理是一种“基于设计的治理”，<sup>[1]</sup>因此除了对算法应用层面的民主控制外，更重要的是关注算法设计层面的“反民主”倾向及其可能引发的算法治理合法性与正当性危机。例如在美国，随着市政算法的广泛应用，人们开始关注其对民主价值的侵害，“这些系统的开发者既未经选举，也无法问责，却被赋予了决定市政实践和优先事项的重大权力”。（[22]，p.136）事实上，由于算法的内在政治性属性，其技术设计本身就是一项涉及政治价值选择的公共决策过程，<sup>[12]</sup>需要体现并促进民主价值。因此，基于对算法设计与民主价值重要性的考量，本文提出一种“参与式设计”的理念，谋求通过算法设计过程的重构来改变不同主体之间权力、信息与价值方面的不对称以及算法公益性、公开性与公平性的异化问题，从而促进算法公共性的实现。

#### 1. 参与式算法设计的基本理念

参与式算法设计理念的提出即是要谋求改变当前算法设计中技术逻辑与商业逻辑主导的态势，将其转变为社会主导或公共价值主导。从算法治理过程来看，针对算法公共性异化风险问题，要将改革的重点由算法应用结果产生之后的监督、审计与问责转变为算法投入应用之前的设计阶段，从源头上防范风险的发生。这种治理重心的“前移”有助于解决因算法的难以解释或相对自主性特征而可能引发的“非预期后果”问题。正如温纳曾指出的，“非故

意的后果并非没有意图,这意味着在初始计划中鲜有任何内容旨在阻止它们发生”。([25], p.84)因此需要在算法设计阶段全方面考虑算法可能的风险来源及其后果,并设计出有效的防范或应对方案。如果对可能的风险后果缺乏有效应对思路的话,则意味着算法设计的“不成熟”或“失败”,此时算法的盲目应用即成为一种不正当的选择。

参与式算法设计强调通过算法设计过程的广泛参与来实现社会整体价值的嵌入。传统的算法研究中已经开始强调“人在回路中”(human-in-the-loop),注重将人类引入算法自动化决策过程中,确保算法系统接受人类的监督与控制。然而这种“人类介入”主要局限在算法技术专家的参与上。参与式算法设计所强调的则是“社会在回路中”(society-in-the-loop),<sup>[26]</sup>谋求通过更广泛的利益主体之间的参与协商达成算法的“社会契约”,以此作为算法设计的基本原则以及算法治理的价值基础。

将人类社会的重要价值与伦理原则嵌入算法的设计,这已经成为许多国家和地区改革中的重要原则。例如欧盟发布的《值得信任的AI伦理准则》中即主张“内嵌伦理价值的设计”(values-by-design)。<sup>[27]</sup>我国发布的《新一代人工智能伦理规范》中也强调要在算法设计等环节中提升透明性、可解释性、可靠性等价值规范。在算法设计价值嵌入的具体实现方式上,有人主张“自上而下”路径,谋求将人类普遍接受的价值原则转化为算法的代码规则;也有人主张“自下而上”路径,主张让人工智能算法自主去发现和学习建构伦理道德原则。<sup>[28]</sup>然而上述路径在本质上体现的仍是专家主导与技术理性的逻辑。参与式算法设计则意在打破专家与技术理性的主导,让社会多元主体及其价值原则和利益诉求成为影响算法设计发展方向的基础性力量。

从参与式算法设计的目标定位来看,通过多元主体、价值与利益之间的充分互动与参与协商,意在全面了解算法设计的各种可能性,进而寻求算法技术发展的替代性方向。算法技术发展的新方向需要能够改变因算法的数据依

赖性、统计相关性以及相对自主性等技术特性而可能引发的负面后果,使算法的技术属性能够更好地与公共性价值相契合。

## 2. 参与式算法设计的实践模式

由于算法的高度复杂性,“外行”很难理解其技术构成与运行原理,这使得算法治理中的实质性参与难以实现,设计层面的广泛参与更是容易停留于一种激进的幻想。不过,无论是在理论层面还是现实实践中,算法设计中参与协商的必要性与可能性已经开始引起人们的关注。通过对既有研究与实践的梳理,下文将从参与主体、范围与实践机制三个方面对参与式算法设计的实践模式进行初步考察。

(1) 算法设计的参与主体。参与式算法设计主张算法设计过程中参与主体的广泛性,除了算法领域的平台企业、技术专家外,包括政府、社会组织、人文社会科学学者乃至普通公众用户在内的算法利益相关者均应当参与其中。其中,由于算法治理中算法系统的公共基础设施定位,政府管理部门应当成为算法基础设施设计和部署中的主导者,在采购招标、合同签订和日常运营过程中基于公共价值原则来平衡算法的商业性或资本性。在一些高风险领域则需要考虑由国家主导设置公共算法平台,以防范特殊利益嵌入算法设计而可能威胁社会乃至国家安全。政府主体的参与还需要涉及一线工作者或街头官僚,他们是算法系统的直接使用,日常工作中积累的算法应用情境知识和执行经验可以成为算法系统优化设计的重要反馈信息。<sup>[29]</sup>

在社会领域,媒体与社会组织的参与十分重要,现实中正是依靠媒体或社会组织的披露才揭示了许多算法应用导致的偏见、歧视等不公平现象。例如美国司法部门中广泛采用的预测罪犯再犯概率的COMPAS系统即是在一个名为ProPublic的非营利组织的调查下才被发现预测结果中存在严重的种族歧视问题。而人文社会科学领域相关学者的参与同样可以为算法设计提供不同的观察视角,特别是关于算法社会政治影响以及风险后果的相关研究应当成为算法设计中的重要考量因素。此外,普通社会公



众在算法设计中的参与作用不可或缺。尽管公众往往缺乏相关专业技术知识，但这并不意味着他们对算法的“无知”。公众在与算法打交道的过程中会基于个人对算法的感知和体验而形成丰富的“算法想象”，<sup>[30]</sup>这些应当成为算法优化设计中的重要输入。当前公众用户的身份主要体现为消极的算法数据提供者，因此需要对公众进行赋权，赋予其数据公民身份，使其个人数据权利得到充分保障，以此作为参与算法正当性评估与优化设计的基础。

(2) 算法设计的参与范围。算法的设计是一个涉及问题界定、目标/价值选择、数据收集与整合、数据标签、模型构建、应用前评估、应用结果反馈与模型更迭等在内的复杂过程，参与式算法设计需要体现到每一个环节中。例如，在问题界定阶段，要通过广泛参与讨论来明确应用算法工具所要解决的公共问题属性，特别重要的是评估该问题是否需要以及能够通过算法工具予以解决，以避免算法工具对其他治理工具的排斥。在治理目标与价值的选择阶段，要通过参与讨论来呈现算法设计中涉及到的目标与价值权衡问题，例如准确性与公平性之间的权衡，以及对“公平”的各种不同界定标准之间的权衡等。同时要注意治理目标或价值向算法代码的“转译”问题，<sup>[31]</sup>能否做到准确的转译将直接决定算法结果是否会偏离预定目标与价值，这需要通过广泛的参与监督来确认。

由于算法的数据依赖性，能否避免数据选择中的历史偏见或统计性偏见等问题十分关键，而通过广泛参与纳入更为多元的资料、价值与观点有助于提升算法训练数据的可靠性与代表性。同时，对初始资料数据的准确标签同样重要，这是算法模型构建的基础。准确的标签需要全面了解资料数据产生的情境特征，以及与算法目标价值相关的人文社会科学知识，这些单纯依靠算法技术专家无法有效获取，需要通过广泛参与来予以弥补。此外，算法模型的设计并非一次性的，而是一个随着应用前评估与应用后效果反馈持续更迭的过程，这一过程同样需要多元主体特别是算法用户的广泛参

与来予以实现。

(3) 算法设计的参与机制。参与式算法设计理念的实现需要具有可行性和操作性的实践机制的支撑。在这方面，许多国家和地区的算法治理实践中已经出现了多种形式的参与机制，并取得了较为明显的参与成效。例如，欧盟发起设立了一个名为“欧盟AI联盟”的在线论坛，该论坛对所有人开放，用来为欧盟人工智能专家委员会收集关于人工智能算法设计的各种意见反馈。<sup>[32]</sup>2017年，由来自麻省理工学院媒体实验室、世界经济论坛等多个机构的倡导者共同发起设立了“开放算法”(Open Algorithms)工程项目，意在增强算法设计中的公民参与，更好地利用大规模人类行为数据来追求公共价值。<sup>[33]</sup>

此外，许多地方性的算法参与实践中则采取了更为直接的协商对话形式，通过直接听取利益相关主体的观点诉求来获取算法设计优化的反馈意见。在这方面，有机构对英国公共服务领域算法决策中的公民参与状况进行了调查总结，发现诸如“公民陪审团”(citizens jury)等参与机制在算法的优化中发挥了重要作用。<sup>[34]</sup>而在我国，外卖算法领域的协商恳谈会已经在上海召开，通过听取外卖员这一算法直接利益相关者的意见诉求来优化算法是协商恳谈的重要内容，<sup>[35]</sup>“算法取中”原则的提出则是通过参与协商实现算法设计思路转变的重要体现。类似的参与机制及其实践经验需要引起算法治理相关部门的关注并进行借鉴。

## 结 语

当前，人工智能算法已经在经济社会领域取得了显著的应用成效，并逐渐成为公共事务治理领域中的重要治理工具，从而推动着“算法治理”的快速发展。不同于算法的商业应用，算法治理场景中的算法应用具有特殊性。一方面，算法治理会直接影响政府提供的公共服务质量以及所有社会公众的切身福祉；另一方面，面对算法的重大影响，社会公众无法选择“退出”而免于受到算法的影响。<sup>[36]</sup>正是由于算法



治理的这种特殊性,如何确保其公共性价值的实现至关重要。

当前学界针对算法治理公共性价值实现中的潜在风险及其治理问题已经展开了许多讨论,然而受到“技术中立性”或“工具中立性”思维的潜在影响,较少有研究者针对算法技术这一治理工具本身与公共性价值之间的内在关联进行系统考察。本文则通过技术政治视角的引入对算法的“政治性”属性进行了揭示。特别是由于算法的“内在”政治性属性,要求我们不仅要关注算法应用过程与结果进行监督、控制与问责,还需要关注算法自身的设计是否与公共性价值相契合的问题。现实来看,作为新兴科技发展的最前沿,算法技术研发设计所体现的主要是技术自身的演化逻辑以及商业资本的逐利倾向。而算法公共性以及参与式算法设计理念的提出则意在促成我们对于算法技术、社会与人类自身关系的新的想象,并由此指引我们对当前的算法秩序以及未来的算法技术发展路径进行变革。只有这样,算法社会才能够成为值得期待的社会愿景,而基于算法的公共治理也才能够更好的实现公共利益或公共价值创造的目标诉求。

#### [参考文献]

- [1] 张敏. 算法治理: 21世纪的公共管理现代化与范式变革[J]. 政治学研究, 2022, (4): 50-62.
- [2] 昌诚、张毅、王启飞. 面向公共价值创造的算法治理与算法规制[J]. 中国行政管理, 2022, (10): 12-20.
- [3] 杨雪敏、刘特、郑跃平. 数字治理领域公私合作研究述评: 实践、议题与展望[J]. 公共管理与政策评论, 2020, 9(5): 3-18.
- [4] 编辑部. 重塑平台的公共性[J]. 文化纵横, 2021, (1): 6-7.
- [5] 郭林生、李小燕. “算法伦理”的价值基础及其建构路径[J]. 自然辩证法通讯, 2020, 42(4): 9-13.
- [6] 张永忠、张宝山. 算法规制的路径创新: 论我国算法审计制度的构建[J]. 电子政务, 2022, (10): 48-61.
- [7] Winner, L. 'Do Artifacts Have Politics?'[J]. *Daedalus*, 1980, 109(1): 121-136.
- [8] 达拉斯·斯迈思. 自行车之后是什么? ——技术的政治与意识形态属性[J]. 王洪喆译, 开放时代, 2014, (4): 95-107.
- [9] 郑崇明. 警惕公共治理中算法影子官僚的风险[J]. 探索与争鸣, 2021, (1): 103-109.
- [10] 林文源. AI的能与不能: 以医疗与照护为例发挥社会想象力[J]. 人文与社会科学简讯, 2019, 20(2): 99-103.
- [11] Stinson, C. 'Algorithms are Not Neutral: Bias in Collaborative Filtering'[J]. *AI and Ethics*, 2022, 2: 763-770.
- [12] Simons, J., Ghosh, D. 'Utilities for Democracy: Why and How the Algorithmic Infrastructure of Facebook and Google Must Be Regulated'[R]. Brookings Institution, 2020.
- [13] 肖林. 协商致“公”——基层协商民主与公共性的重建[J]. 江苏行政学院学报, 2017, (4): 104-113.
- [14] 苗红培. 政府向社会组织购买公共服务的公共性保障研究[D]. 济南: 山东大学, 2016.
- [15] 王明杰、李晓月、王毅. 西方学界公共性理论研究述评及展望[J]. 公共管理与政策评论, 2021, 10(4): 155-168.
- [16] 伊格纳斯·卡尔波卡斯. 算法治理: 后人类时代的政治与法律[M]. 邱遥堃译, 上海: 上海人民出版社, 2022.
- [17] Hartmann, K., Wenzelburger, G. 'Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA'[J]. *Policy Sciences*, 2021, 54: 269-287.
- [18] Bucher, T. 'Neither Black Nor Box: Ways of Knowing Algorithms'[A], Kubitschko, S., Kaun, A. (Eds.) *Innovative Methods in Media and Communication Research*[C], Cham: Palgrave, 2016, 81-98.
- [19] Burrell, J. 'How the Machine “Thinks”': Understanding Opacity in Machine Learning Algorithms'[J]. *Big Data & Society*, 2016, 1: 1-12.
- [20] Coglianese, C., Lampmann, C. 'Contracting for Algorithmic Accountability'[J]. *Faculty Scholarship at Penn Law*, 2021, 1: 175-199.
- [21] Magrani, E. 'New Perspectives on Ethics and the Laws of Artificial Intelligence'[J]. *Internet Policy Review*, 2019, 8(3): 1-19.
- [22] 本·格林. 足够智慧的城市: 恰当科技与城市未来[M]. 李丽梅译, 上海: 上海交通大学出版社, 2020.
- [23] Busuioc, M. 'Accountable Artificial Intelligence: Holding Algorithms to Account'[J]. *Public Administration Review*, 2020, 81(5): 825-836.
- [24] 刘永谋、谭泰成. 论当代技术哲学的政治转向[J]. 理论探索, 2021, (6): 51-57.

- [25] 兰登·温纳. 自主性技术: 作为政治思想主题的失控技术 [M]. 杨海燕译, 北京: 北京大学出版社, 2014, 280.
- [26] Rahwan, I. 'Society-in-the-Loop: Programming the Algorithmic Social Contract' [J]. *Ethics and Information Technology*, 2018, 20(1): 3–11.
- [27] AI HLEG. 'Ethics Guidelines for Trustworthy AI' [R]. The European Commission Report, 2019.
- [28] 齐琳琿. “有道德的算法”: 重构智媒时代公共性 [J]. 青年记者, 2021, (24): 32–34.
- [29] Vogl, T., Seidelin, C., Ganesh, B., et al. 'Smart Technology and the Emergence of Algorithmic Bureaucracy: Artificial Intelligence in UK Local Authorities' [J]. *Public Administration Review*, 2020, 80(6): 946–961.
- [30] Bucher, T. 'The Algorithmic Imaginary: Exploring the Ordinary Affects of Facebook Algorithms' [J]. *Information, Communication & Society*, 2017, 20(1): 1–15.
- [31] Loi, M., Ferrario, A., Viganò, E. 'Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms' [J]. *Ethics and Information Technology*, 2021, 23(3): 15–27.
- [32] Sloane, M. 'Here's What's Missing in the Quest to Make AI Fair' [J]. *Nature*, 2022, 605: 9.
- [33] Lepri, B., Oliver, N., Letouze, E., et al. 'Fair, Transparent, and Accountable Algorithmic Decision-making Processes' [J]. *Philosophy & Technology*, 2018, 31: 611–627.
- [34] Warne, H., Dencik, L., Hintz, A. 'Advancing Civic Participation in Algorithmic Decision-making: A Guidebook for the Public Sector' [R]. Data Justice Lab, 2021.
- [35] 周楠. 上海首份外卖骑手民主协商纪要诞生 [N]. 解放日报, 2022-07-10 (3).
- [36] Wenzelburger, G., König, P., Felfeli, J., et al. 'Algorithms in the Public Sector: Why Context Matters' [J]. *Public Administration*, 2022, 11: 1–21.

[责任编辑 李斌]